International Journal of Advanced Information in Engineering and Technology (IJAIET)
Vol.1, No.1, September 2014

Analysis of Babble Noise based on MFCC in Stuttered Speech

Miss. Salma Jabeen, Asst. professor,
Dept. of Computer Science and Engg.,
Ghousia College of Engineering,
Ramanagara,India.
salmakhayum@gmail.com

Dr. K.M. RaviKumar, professor & Head,
Dept. of Electronics & Communication Engg.,
S. J. C. Institute of Technology,
Chickballapur-562101,India..
kmravikumar@rediffmail.com

***Abstract*— This paper describes the effect of 0dB and 10dB babble noise on stuttered speech. 100 samples are collected from the subjects (stutterer), among which 80 samples are used to prepare database dictionary and 20 samples for testing. The samples are segmented and pass through Mel Frequency Cepstral Coefficient (MFCC) feature extraction and statistical estimation to prepare dictionary, then testing samples compared with the database. The results obtained are 100%. When the same testing samples are added with the babble noise of 0dB and 10dB, the result degrades; the degraded results are enhanced by using sparse matrix enhancement technique. The results obtained after enhancement are 60-95% depending on the samples used for analysis.**

Keywords-- Stutter, subject, noise, MFCC.

I. INTRODUCTION

Stuttering is a complex disorder involving interactions among what the child does, how he feels, and what he thinks. The child might have breaks in the forward flow of speech, such as repeating a sound or syllable, Prolongation, or being unable to say a word at all. The child may begin to avoid and fear speaking; express frustration at being unable to talk; or use other behaviors to help get speech affecting, for example blinking his eyes, nodding his head, or stamp his foot. The child might express his thoughts through questions or comments such as "Why can't I talk?" or "My mouth is broken." Not all children will exhibit negative feelings or thoughts about their speech. As the stuttering progresses, the likelihood of developing negative attitudes toward communication increases.

Stuttering is also known as stammering in the United Kingdom. Stammering is often associated with “Repetitions”. Part-word-repetition / syllabic repetitions are one of the defining elements of early stammering. The dominant features of speech reported are: (1) Word Repetitions, but not part-word-repetition is a prevalent feature of early stammering. (2) In early stammering, there is a high proportion of repetition in general, as compared to other types of disfluency like prolongation. Stuttering is also called as stammering or disfluent speech, in speech disorder. This is different when compare with normal repetition of words that children may do when learning to speak. Normal developmental stuttering / disfluent may occur when the child is between the ages of 18 months to 5 years. This may include repeating words / phrases, pronunciation of words, leaving out words or sounds. There are different types of stuttering, that include the following: 1) Developmental Stuttering 2) Neurogenic stuttering 3) psychogenic stuttering. Although stuttering may cause emotional problems, but it cannot believed to be the result of emotional problems. In this paper we consider repetition as the major characteristics feature of stuttering for studying the effect of 0db and 10db babble noise on stutterer.

Noise may be defined as any unwanted signal that interferes with the communication or processing of an information bearing signal [12].

Babble noise is also called Speech babble. The source of this babble is 100 people speaking in a canteen. The room radius is over 2 meters; therefore, individual voice is slightly audible. The sound intensity during the recording process was 88 dB [16].

II. DATA DESCRIPTION

100 Stuttered speech samples are collected in which 80 are used to prepare database dictionary and 20 for testing. Processing speech is often difficult for speech recognizers and the stuttered speech will be more difficult. In this work the subjects were made to read a given text aloud. Recognizing read speech is easier for recognition system because the system has the prior knowledge about the subjects intended words. Although read speech is not typically as rich in disfluencies as spontaneous speech, the stutters are often disfluent when reading aloud [13,14]. The read speech task is used as a part of most stuttering assessments [14,15]. The speech samples were recorded by making them to read English “All Phoneme passage”. In this passage some of the possible part word and whole word repetition are as follows

TABLE I
PART WORDS AND WHOLE WORDS

Part words	**Whole words**
gra	Grand
dra	Dresses
co	Coat
ba	Banana

III. METHODOLOGY

A. *Database Dictionary*

Out of 100 samples collected, 80% of it i,e. 80 samples are used to prepare database dictionary by calculating MFCC of each sample. For these MFCC coefficients, the statistical properties are applied to obtain the reduce coefficients and classifying the coefficients to get distance metric and parameters. A block diagram is show in fig. 1.

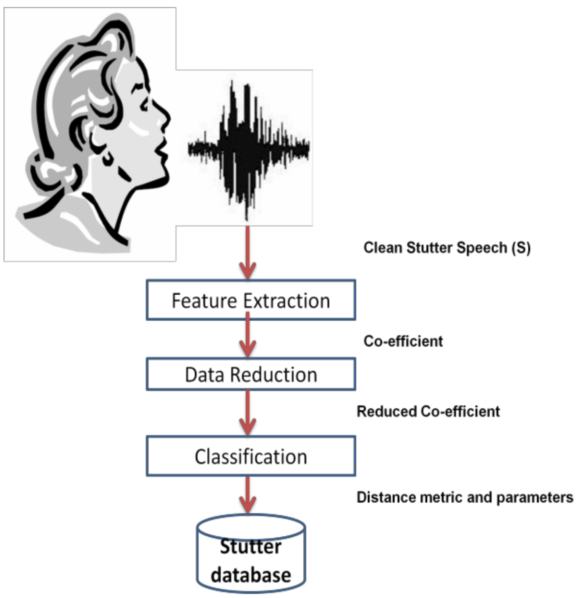

Fig. 1 Block diagram of Database dictionary words

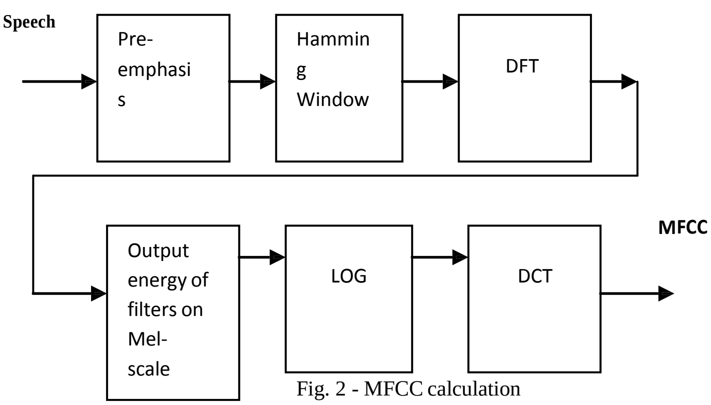

Fig. 2 - MFCC calculation

Step by step procedure to calculate MFCC

Mel Frequency Cepstrum Coefficients

The block diagrams for calculating MFCCs are as shown in Fig. 2.

Step 1: Pre emphasis for the complete sound file.
The transfer function for the pre-emphasis is given by

Where α is a constant, which has a typical value of 0.97.

Step 2: Framing the complete sound file to get many blocks
After pre-emphasis, the speech signal is subdivided into frames. This process is the same as multiplying the entire speech sequence by a windowing function.

where F[n] is the entire speech sequence, F_m [n] is a windowed speech frame at time m and W[n] is the windowing function.

Step 3: Hamming windowing for each frame
One of the most commonly used windows is the Hamming window.

(3)

In the above equation, N is the length of the windowing function.

Step 4: Fast Fourier Transform for each speech frame
FFT will be applied to each speech frame in order to obtain its speech spectrum.
S_{fft}[k], for $0 \leq k \leq N-1$, denotes the N-point FFT spectrum of one Hamming-windowed speech frame.

Step 5: Mel Filter Bank Processing for each frame
The Mel frequency filter bank is a series of triangular bandpass filters, which mimics the human auditory system. The filter bank is based on a non-linear frequency scale called the Mel scale. The Mel frequency scale is a psychoacoustic measure of pitches judged by human. According to Stevens et al., a 1000Hz tone, with 40dB above the listener's threshold, is defined as having a pitch of 1000 mels. Below 1000Hz, the Mel scale is approximately linear to the linear frequency scale. The following equation describes the mathematical relationship between the Mel scale and the linear frequency scale,

Where Mel_{frq} is the Mel frequency in mels and f is the linear frequency in HZ
In order to model the perceived loudness of a given signal intensity, the filter outputs are compressed by a logarithmic function Eq. (5)

In the above equation, $C_{m(ln)}$ is the logarithmically-compressed output of the m^{th} filter.

Step 6: Discrete Cosine Transform for each frame
DCT is applied to the filter outputs and the first few coefficients are grouped together as a feature vector of a particular speech frame.
Suppose p is the order of the Mel scale cepstrum. The feature vector is obtained by considering the first p DCT coefficients. Mathematically, the i^{th} MFCC coefficient can be expressed by the following formula.

Similarly after adding babble noise to the clean stutter speech and the noisy speech enhancement is done by using sparse matrix technique, a block diagrams is show in fig. 3 and fig. 4

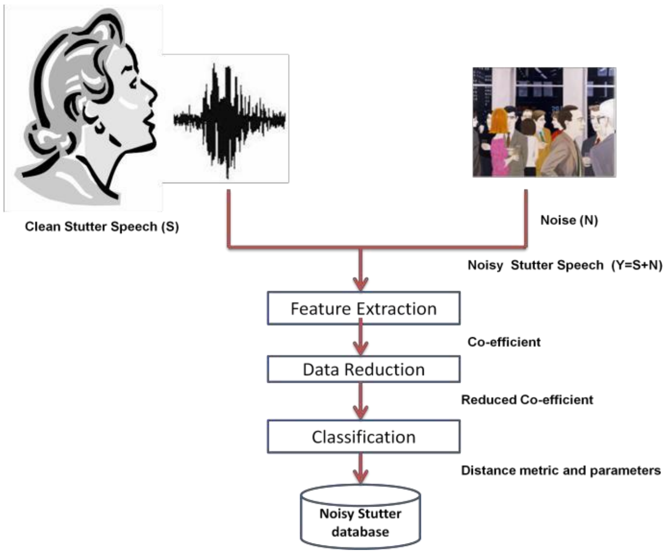

Fig. 3 Block diagram of Stutter speech with babble noise

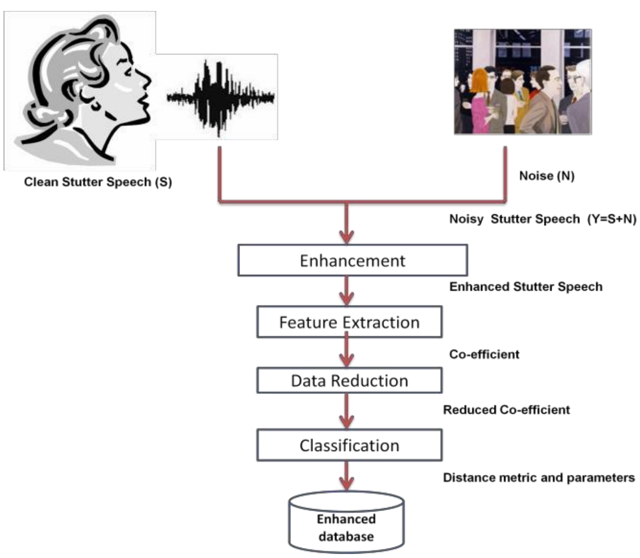

Fig. 4 Block diagram of Noisy Stutter speech with enhancement

B. Test and Classify

20 samples are select for testing; it will be passed one by one to MFCC and then statistical property and stored in a vector output1. As shown in fig.5.

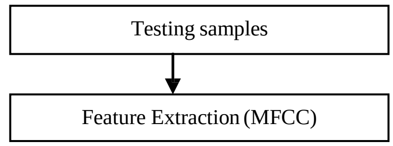

Fig. 5 Flow chart of testing data

Similarly the noisy and enhanced samples testing flow chart is shown in fig.6 and fig.7

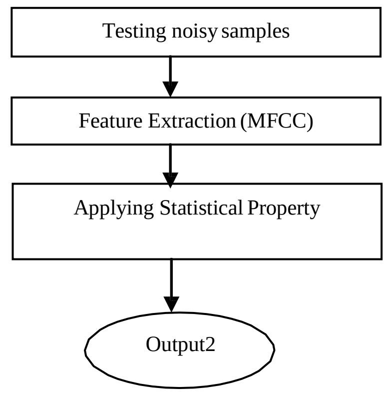

Fig. 6 Flow chart of testing data with babble noise

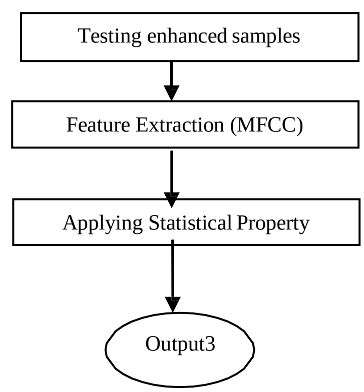

Fig. 7 Flow chart of testing enhanced samples

Comparing output1 with stutter database, output2 with noisy stutter database and output3 with enhanced database using Euclidean distance for KNN classification.

K-nearest neighbor (k-NN)

The algorithm to compute the K-nearest neighbors is as follows:

1) Determine the parameter K, number of nearest neighbors. This value is all up to you.
2) Calculate the distance between the class label and all the samples from dataase using Euclidean distance, which gives Pair wise distance between two sets of observations.
   The expression for Euclidean Distance measuring is given by Eq. (7).

   Where a1….an are the training Set with the class label b1….bn.
3) Sort the distances for all the samples and determine the nearest neighbor based on the K-th minimum distance.
4) Get all the Categories of data's for the sorted values which fall under K.
5) Use the majority of nearest neighbors as the calculation value.

IV. RESULTS

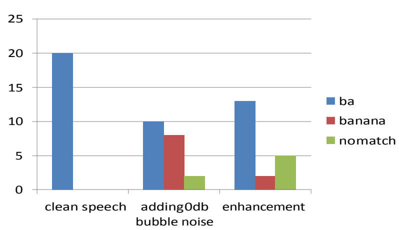

Fig 8(a): ba and banana with 0db bubble noise & enhancement

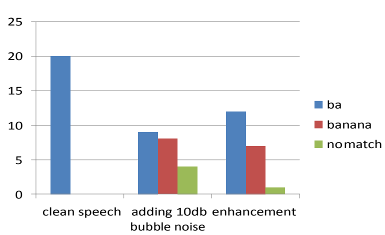

Fig 8(b): ba and banana with 10db babble noise & enhancement

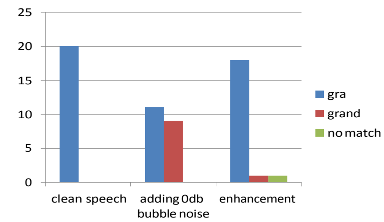

Fig 8(c): gra and grand with 0db bubble noise & enhancement

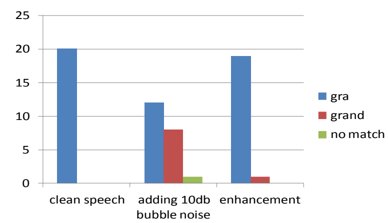

Fig 8(d): gra and grand with 10db babble noise & enhancement

Fig 8 shows the graphical representation of results of fig 5,6 and 7. In fig 8(a), 8(c) represents the results of part word and whole word with 0db babble noise. In fig 8(b), 8(d) represents the results of part word and whole word with 10db bubble noise. The blue bar represents the part word results, the red bar represents the whole word results, the green bar represents the no match results.

Similarly the results in percentage are shown in table II.

TABLE II

RESULTS IN PERCENTAGE

Part words	% of result for clean speech	% of result with 0db babble noise	% of Enhancement from noisy result(0db)	% of result with 10db babble noise	% of Enhancement from noisy result(10db)
ba	100%	50%	65%	45%	60%
gra	100%	55%	90%	60%	95%
co	100%	70%	90%	60%	75%
dra	100%	50%	65%	60%	70%

From table II it can be observed that, the recognition result is 100% for clean speech, irrespective of sound repeated. When noise is added the percentage reduces to the range of 45% to 70%, depending on the sound and the type of noise added. Similarly after enhancement the result is improved towards 60% to 95% depending on the sound used.

V. FUTURE WORK

The number of samples can be increase and the variation of result may be analyzed.

The effect of other standard noise with different dB of SNR can be analyzed.

VI. CONCLUSION

In this paper we analyze the effect of 0dB and 10dB babble noise effect on stutter speech such as repetition, and Enhancement of noisy samples by degrading the noise effect to some extent. The present work will help the clinician to objectively assess the stutterer in noisy environment such as 100 people speaking in a canteen.

Although different Altered Auditory Feedback systems are available for the treatment of adult stutter [7], there is a need for better recognition and identification of the source of stuttering. Many methods are proposed for the recognition and objective assessment of stuttered speech [7 - 10], yet there is a wide scope for improvement.

REFERENCES

[1] Bloodstein,O.(1970).Stuttering and normal disfluency a continuity hypothesis . British Journal of Disorders of communication,5,30-39 W.Johnson et al., "The onset of stuttering, minneapolies :University of Minnesata press," 1959.

[2] E.Yairi and B.Lewis, "Disfluencies at the onset of stuttering," Journal of speech & Hearing Research, vol.27, pp. 154-159, 1984.

[3] E.G.Conture, "Stuttering" Englewood cliffs, New Jersey: Prentice-Hall, 2nd edition, 1990.

[4] http://medicine.yale.edu/

[5] K.M. Ravikumar, Balakrishna Reddy, R. Rajagopal and H.C. Nagaraj, "Automatic Detection of Syllable Repetition in Read Speech for Objective Assessment of Stuttered Disfluencies", 2008.

[6] Fuhai, L., Jinwen, M., Dezhi, H., MFCC and SVM Based Recognition of Chinese Vowels. Lecture Notes in Artificial Intelligence, 3802:812–819, 2005.

[7] K.M.Ravikumar, and Dr. R.Rajagopal "Altered Auditory Feedback Systems for Adult Stutter", Proceedings of the Sonata international Conference on Computer Communication and Control, November 2006, pp. 193-196.

[8] K.M.Ravikumar , Sachin Kudva , Dr. R.Rajagopal and Dr.H.C.Nagaraj, "Development of a Procedure for the Automatic Recognition of Disfluencies in the Speech of People who Stutter", International Conference on Advanced Computing Technologies, Hyderabad, India, December 2008, pp. 514-519.

[9] K.M.Ravikumar , S.Mahadev , Dr. R.Rajagopal and Dr.H.C.Nagaraj, "An Algorithm to Compute the number of Repetitions and its Iterations from Stuttered Speech for Objective Assessment of Stuttering, ICOICT, International Conference on Optoelectronics", Information and Communication Technologies, Trivandrum, Kerala, India, February 2009, pp. 147-150.

[10] K.M.Ravikumar , T. Satish , Dr. R.Rajagopal and Dr.H.C.Nagaraj, "Bayseian Classifier for Classification of Normal Nonfluency and Fluent Speech from Stuttered Disfluencies", ICOICT, International Conference on Optoelectronics, Information and Communication Technologies, Trivandrum, Kerala, India, February 2009, pp. 202-205.

[11] K.M.Ravikumar , Dr. R.Rajagopal and Dr.H.C.Nagaraj, "An Approach for Objective Assessment of Stuttered Speech Using MFCC Features", ICGST, International Journal on Digital Signal Processing, Vol.9, June 2009, Issue.1, pp. 19-24.

[12] "Advanced Digital Signal Processing and Noise Reduction", Second Edition. Saeed V. Vaseghi Copyright © 2000 John Wiley & Sons Ltd ISBNs: 0-471-62692-9 (Hardback): 0-470-84162-1 (Electronic).

[13] Baneree. S, Beck J.E and Mostow. J ,"Evaluating the effect of predicting oral reading miscues" in proceedings of Eurospeech, Septemer 2003a,Geneva,Switzerland.

[14] G. Riley, "Stuttering severity instrument for children and adults", third. Austin,TX,Pro-Ed,1994.

[15] H.H Gregory, J. H. Campbell, C. Gregory and G. Hill, "Stuttering therapy: Rationale and Procedures", Boston, Pearson Allyn and Bacon, 2003.

[16] Yuxuan Wang and DeLiang Wang "feature denoising for speech separation in unknown noisy environments "ICASSP 2013, 978-1-4799-0356-6 ©2013 IEEE.

[17] http://spib.rice.edu/spib/data/signals/noise/

Authors Biography

Miss. Salma Jabeen, Asst. professor, CSE department, GCE, RMGM. She has completed her M.Tech in the field of Software Engineering in the year 2009 (VTU, Belgaum). She has nine years of teaching experience and has published one research papers in National Conference, one in International Conference and journal.

Dr. K.M. RaviKumar, professor & Head, Dept. of Electronics & Communication Engg., S. J. C. Institute of Technology, Chickballapur. He has completed his M.Tech in the field of Biomedical Instrumentation in the year 2002 & completed PhD in 2011 (VTU, Belgaum). He has 15 years of teaching experience and has published ten research papers in International Conferences, six research papers in National Conference and 10 in International journals. He holds membership of various professional societies which include ISTE, SIS, MIE and BMESI. His field of interest includes DSP and Communication Systems.